DOCUMENT RESUME

ED 333 309 CG 023 464

AUTHOR Alston, Reginald J.; Mngadi, Simphiwe
TITLE Vocational Aptitude Testing in Counseling and Rehabilitation: Historical Developments and Application Issues. Kentucky Studies in Rehabilitation.
INSTITUTION Kentucky Univ., Lexington. Graduate Program in Rehabilitation Counseling.
PUB DATE 90
NOTE 64p.
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Aptitude Tests; Career Counseling; Counseling Techniques; *Disabilities; History; *Human Services; Intelligence; *Professional Personnel; *Rehabilitation Counseling; *Test Use; Vocational Evaluation

ABSTRACT

An examination of the continuing debate among human service professionals about a precise definition of aptitude, with special attention devoted to the relationship of interests and intelligence to aptitude, is provided in this monograph. An historical look at aptitude test development is provided with a discussion of how individuals and events influenced the evolution of the tests. The General Aptitude Test Battery, a commonly used multiple aptitude test in rehabilitation, is analyzed to explain how multiple aptitude vocational tests were developed to provide a comprehensive evaluation of vocational potential. In addition, the practical utility of vocational aptitude tests in human services is addressed and the controversy regarding the use of general cognitive ability measures as opposed to specific aptitude tests to collect data and make placement decisions is examined. Finally, a discussion of the application issues involved in testing persons with disabilities is provided. (BHK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED338809

Kentucky Studies in Rehabilitation

Vocational Aptitude Testing in Counseling and Rehabilitation: Historical Developments and Application Issues

by
Reginald J. Alston
and
Simphiwe Mngadi

Edited by:
Ralph M. Crystal

Graduate Program in Rehabilitation Counseling
Graduate School/College of Education
124 Taylor Education Building
University of Kentucky
Lexington, Kentucky 40506
(606) 257-3834
FAX (606) 257-3835

1990

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Reginald J. Alston

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

CG023464

Vocational Aptitude Testing in Counseling and Rehabilitation: Historical Developments and Application Issues

by

Reginald J. Alston
and
P. Simphiwe Mngadi

Edited by:
Ralph M. Crystal

Graduate Program in Rehabilitation Counseling
124 Taylor Education Building
University of Kentucky
Lexington, Kentucky 40506
(606) 257-3834

Copyright © 1990 by the University of Kentucky Research Foundation

TABLE OF CONTENTS

FOREWORD		iii
CHAPTER I	HOW BEST TO DEFINE APTITUDE	1
	The Relationship Between Interests and Aptitudes	3
	The Influence of Intelligence on Aptitudes	5
CHAPTER II	MAJOR CONTRIBUTORS AND HISTORICAL EVENTS IN THE EVOLUTION OF APTITUDE TESTING	9
	The Chinese Contribution	9
	European and American Contributions	10
	Aptitude Testing and World War I	20
	Early Tests of Musical and Artistic Aptitudes	23
	Clerical Aptitude Testing	24
	Mechanical and Manual Aptitude Testing	25
	Spatial Aptitude Testing	26
CHAPTER III	THE DEVELOPMENT OF MULTIPLE APTITUDE TESTS	27
	Utilization of Factor Analysis	28
	The General Aptitude Test Battery (GATB)	30
CHAPTER IV	ISSUES INVOLVED IN USING APTITUDE TESTS TO PREDICT VOCATIONAL POTENTIAL	33
	Practical Utility of Aptitude Tests	34
CHAPTER V	APPLICATION ISSUES INVOLVED IN TESTING PERSONS WITH DISABILITIES	39
	Visual Disabilities	42
	Hearing Disabilities	43
	Cognitive Deficiencies	43
	Physical Disabilities	45

General Test Modifications 45

Client Perspective 47

CHAPTER VI SUMMARY AND CONCLUSION 49

FOREWORD

It can be debated intelligently that vocational aptitude testing is the key to successful rehabilitation of persons with disability. If the primary responsibility of the rehabilitation professional is placement of the client in employment, and an assessment of the person's vocational potential is necessary to make the placement, then it is logical to suggest that a quality vocational aptitude evaluation is paramount. If you agree with the statements above, then what exactly are aptitudes and what exactly is vocational aptitude testing? There is mention of vocational aptitudes throughout the rehabilitation literature. However, there is a scarcity of information which defines aptitude and its significance to vocational counseling in rehabilitation. Also, there is an absence of material which chronicles the development of vocational aptitude tests. This monograph was developed to fill that void in the rehabilitation literature.

The curriculum of rehabilitation counselor education programs that are accredited by the Council on Rehabilitation Education (CORE) contain a course on vocational evaluation. Understanding aptitude testing is an important component of the course. This monograph was designed to be used as an instructional supplement in the vocational evaluation course offered by rehabilitation counselor education programs. It was developed with the belief that students' appreciation for vocational aptitude tests would be enhanced if they had an in-depth understanding of how multiple aptitude tests were

developed and why they are useful in rehabilitation to assess vocational potential.

This monograph contains an examination of the continuing debate among human service professionals about a precise definition of aptitude, with special attention devoted to the relationship of interests and intelligence to aptitude. A historical look at aptitude test development is provided with a discussion of how individuals and events influenced the evolution of the tests. The General Aptitude Test Battery (GABT), a commonly used multiple aptitude test in rehabilitation, is analyzed to explain how multiple aptitude vocational tests are developed to provide a comprehensive evaluation of vocational potential. In addition, the monograph looks at the practical utility of vocational aptitude tests in human services and examines the controversy regarding the use of general cognitive ability measures as opposed to specific aptitude tests to collect data and make placement decisions. Finally, a discussion of the application issues involved in testing persons with disabilities is provided.

Reginald J. Alston

Lexington, Kentucky
September 1990

CHAPTER I

HOW BEST TO DEFINE APTITUDES

Professionals in occupations, such as rehabilitation counseling, which entail providing assistance to individuals who are making educational and career plans and choices, know that the degree to which they are effective depends on their awareness of clients' aptitudes. In everyday language, we often talk of some people having an aptitude for one thing or the other. For instance, people who can manipulate relatively complex quantitative data faster and more accurately than most other people are said to have an aptitude for numbers, and those who seem to learn linguistics with greater ease and speed to have an aptitude for languages. Instruments that purport to measure aptitudes often include subtests designed to measure attributes such as numerical reasoning, manual dexterity, spatial perception, and motor coordination. But what exactly are aptitudes?

In the Handbook for Analyzing Jobs (United States Department of Labor, 1972), aptitudes are defined as "the specific capacities or abilities required of an individual in order to facilitate the learning of some task or job duty" (p. 233). Warren (1934), defined aptitude as "a condition or set of characteristics regarded as symptomatic of an individual's ability to acquire with training some (usually specified) knowledge, skill, or set of responses such as the ability to speak a language, to produce music, etc." (p. 18). These two definitions suggest that though aptitudes are future-oriented, in that they are predictive of future learning potential, they are at the same time indicative of proficiences that individuals already

possess at the time of assessment. In other words, the assumption is that what individuals can do now is indicative of what they will be able to learn to do in the future. This is most obvious in the distinction between ability and aptitude advanced by Woliver and Saeks (1986). According to Woliver and Saeks, "Ability has the implication that a given task can be performed at the present time, whereas aptitude implies that the person can be trained to perform a specific task at some future time."

An interesting note in these definitions is that the U.S. Department of Labor (DOL) stressed the "specificity" of the abilities that indicate aptitudes, whereas Warren (1934) and Woliver & Saeks (1986) emphasized the "specificity" of the tasks that individuals can learn to perform (as predicted from aptitude test performance). Nonetheless, all three definitions do suggest that aptitude does not refer to a simple unitary capacity but rather to a complex, and often interrelated group of abilities. For example, an individual with an aptitude for music has to be able to discriminate between sounds of different pitch and intensity as well as have a good memory and ability to perceive rhythm. Perhaps this multidimensional nature of aptitude is best reflected in Barclay's (1990) definition which refers to aptitudes as "broad clusters of integrated traits and skills reflective of temperament, cognition, and acquired learning" (p. 34).

Bingham (1937) in his delineation of aptitude indicated that though we often talk of an aptitude as a possession or entity that a

person either has or lacks, this is not really the case. Rather than being a possession, aptitude "refers to those qualities characterizing a person's way of behavior which serve to indicate how well he can learn to meet and solve certain specified kinds of problems." In other words, an aptitude is not "extraneous to" an individual, nor is it "possessed" by an individual, but "it is an integral aspect of him as a person ... an attribute of what he is" (Bingham, 1937,p. 21).

As is obvious from these examples, most definitions of aptitudes do not explicitly address the issue of how aptitudes develop. Whereas some scholars alluded to their inheritability (e.g. Bingham, 1937; Carroll, 1985), others like Barclay (1990) believed aptitudes are largely acquired through learning. Despite the lack of conclusive literature on aptitude development, scholars have attempted to demonstrate how interest and intelligence affect aptitudes.

The Relationship Between Interests and Aptitudes

According to Thorndike's law of effect (McConnell, 1986), individuals are more likely to persist in those activities or endeavors from which they derive the greatest satisfaction. This is basically a question of interest and motivation. Bingham (1937) defined interest as "a tendency to become absorbed in an experience and to continue it" (p. 62). He further asserted that interest in an object or activity reveals itself in a concentration of attention to it and persistence in its pursuit. But what do interests have to do with aptitudes? Obviously, interest in an activity (as reflected

in the intensity and duration of attention and persistence) will sharpen the skills involved in carrying out the activity, just as continued attention to an object will lead to a deeper understanding of it and its "interactions" with other objects in its vicinity. This would suggest that aptitudes and interests feed on each other. As Bingham put it, "More often than not, a person can learn to do best what interests him most, and conversely, is likely to develop an interest in doing what he finds he can do best" (Bingham, 1937, p.61).

However, Myers (1941) indicated that there is not always a positive correlation between aptitudes and interests. Sometimes individuals express an interest in fields for which they have low aptitudes, or vice versa - they show little interest in areas for which they exhibit high aptitudes. For example, an individual might be interested in pursuing a career as a surgeon despite not having good dexterity and coordination that are indespensible to a surgeon. Therefore, what is important is not whether an individual expresses an interest in a particular occupation, but whether the person possesses "a general pattern of interests that characterize successful workers in a specific occupation" (Myers, 1941, p.206). It is this "general pattern of interests", likes and dislikes, together with aptitudes that should guide vocational choice.

Bingham (1937) also noted that immaturity and lack of knowledge and/or hands-on experience with the actual activities involved in an occupation might depress whatever correlation there could be between interests and aptitudes. Therefore, only after individuals have (a)

gathered enough information on an occupation, (b) have had an opportunity to observe workers in the preferred occupation at work, and (c) have themselves had a chance to participate in the occupation can their declarations of interest in it be regarded as reliable predictors of future competence and satisfaction in that occupation.

Some professionals express concern about the instability of interests for them to be reliable indicators of future accomplishment (Bingham, 1937; Myers, 1941). Don't we all recall activities that we truly used to enjoy but which we now find abhorrent? Nonetheless, individual interests seldom change completely. For instance, it is not uncommon to hear of individuals who were once enthralled by football, but who now love basketball, or those who were fascinated by building model ships but who now build model planes, or devout readers turning into devout writers. The point is, even if specific interests, per se, change, general interests seldom do. People who are musically, mechanically, verbally, or physically inclined will often stay that way because it is similar kinds of skills or abilities and activity features that will perpetuate an interest in a certain category of activities.

The Influence of Intelligence on Aptitudes

In addition to interests, one other factor that a discussion of aptitudes cannot exclude is that of intelligence. The definitions of aptitudes we have thus far put forth suggest that aptitudes will determine the extent to which individuals will be able to learn to perform certain specific tasks. Barclay (1990) pointed out that aptitudes are "reflective of ... cognition ...." among other things.

This implies that general cognitive ability (intelligence) plays a significant role in specific task performance.

Intelligence is undeniably an elusive construct. It is not surprising, therefore, that debate on its definition is as controversial today as it was six decades ago (Spearman, 1927). In most cases the differences in definition seem to be basically a question of whether intelligence is viewed as a single general ability that pervades all of an individual's functioning or as a composite of numerous specific abilities. One definition that has gained some popularity is the one advanced by Wechsler (Matarazzo, 1978) who defined intelligence as the ability to act purposefully, think logically, and deal effectively with one's environment.

Bingham (1937) viewed intelligence as an individual's ability to solve new problems. He asserted there were three factors involved in the definition: (a) the level of difficulty of the problems one can solve, (b) the range of problems one can solve at this level, and (c) the speed with which he can solve them. Therefore, what Bingham is suggesting basically is "show me a good problem-solver; I'll show you an intelligent person." Of course, problem-solving entails numerous higher order cognitive skills such as abstract thinking, making inferences from words, symbols, and events, retaining them in memory, recognizing interrelationships among them, analyzing and synthesizing, developing hypotheses, testing the hypotheses mentally, and visualizing outcomes (Bingham, 1937).

To contribute to the quest to understand human cognitive functioning, Spearman (1927) employed factor analysis and came up

with a two-factor theory of intelligence - the general (g) and specific (s) factors. Spearman's factor analysis led him to conclude that the high correlations often observed in mental task performance result from the influence of the 'g'-factor, while the 's'-factor affects performance in specific tasks. In other words, in every task performance there is a contribution of 'g' and 's'. Whereas 'g' is shared by all tasks, 's' is task-specific.

Bingham (1937), in agreement with Spearman (1927), also asserted that in estimating a person's aptitudes, one should ascertain "the quality of intelligence" evidenced in solving problems in various areas because individuals differ in the kinds of problems they solve well. Whereas some will excel in mechanical, mathematical, geometric, linguistic or interpersonal problem-solving, others will not. It appears what Bingham and Spearman were suggesting is that though general cognitive ability permeates our general level of functioning, individuals with comparable general cognitive abilities might still differ in terms of special abilities or aptitudes for specific task performance.

Though Thurstone (1938) also agreed with Spearman in that intelligence is not just one entity, he refuted the notion of general intelligence. His own factor analysis led him to conclude that there are seven, and not two, relatively independent intelligence factors. Guilford (1967) proposed even more intelligence factors - at least 120.

Sternberg (1979) advanced yet another theory of intelligence - the information-processing approach. Sternburg's approach emphasizes

the actual processes involved in mental activity, the speed with which information is processed, as well as the strategies applied in problem-solving. The results of a study by Vernon and Jensen (1983), for instance, suggest a positive relationship among the speed of information-processing (as measured through performance in cognitive tasks), general intelligence, and scholastic achievement.

Obviously, the question of intelligence is complex and is therefore unlikely to be resolved in the near future. Nonetheless, psychologists and other professionals continue to study the construct, and results of their studies continue to deepen our insights and broaden our understanding. For instance, a special issue of the Journal of Vocational Behavior (Vol. 29, 1986) was devoted to the role that Spearman's 'g', or general cognitive ability, plays in job performance. This concept will be discussed later in the monograph.

CHAPTER II

MAJOR CONTRIBUTORS AND HISTORICAL EVENTS IN THE EVOLUTION OF APTITUDE TESTING

The preceeding chapter reviewed definitions of aptitude and discussed how the construct relates to interests and intelligence - factors often included in aptitude batteries. In this chapter, the history of aptitude testing will be discussed, starting with the Chinese contribution, proceeding to developments in intelligence testing, and concluding with the development of aptitude testing both during and after World War I.

The Chinese Contribution

Testing in order to determine eligibility for a job has its origins in the Chinese tradition. For example, Chinese government posts were not obtained by nepotism. As far back as 2200 B.C., those who aspired to holding positions in the civil service had to take examinations and demonstrate proficiency in a number of areas. Records of tests used around 1115 B.C show that competence in music, horsemanship, archery, writing, arithmetic, and familiarity with "the rites and ceremonies of public and private life" (DuBois, 1970, p.3) was imperative. Proficiency in these skills was viewed as being predictive of success in public office. During Confucius' time (551 - 478 B.C.), adherence to Confucian moral standards was added as a requirement for candidates for public office, and by 165 B.C. they had to demonstrate their knowledge of the geography of the Chinese empire, civil law, military matters, agriculture, and the administration of revenue (Allen & Yen, 1979). Though employment

examinations underwent many changes over the years, by 1370 A.D. they had somewhat stabilized.

The examinations were administered at three levels: the district level, the provincial level, and the national level. Administration procedures were uniform at all centers, and more than one evaluator graded the examinations to ensure objectivity. Success at each level qualified a candidate for testing at the next. However, only about 3% of the candidates made it to public service (DuBois, 1970).

European and American Contributions

DuBois (1970) stated that when Europeans were first exposed to the Chinese system, they were quite impressed with it and used it as a model for their own civil service examinations. France, for example, introduced employment tests in 1791. England first used similar examinations in 1833 to select civil servants that were to work in India (then a British colony).

In the United States, bills to establish examinations for civil service employees were introduced in Congress in 1860. It was the responsibility of the Civil Service Commission to construct and administer the tests. The items were sampled from a variety of tasks and activities required of individuals in particular occupations. Later, 15% of the items demanded knowledge of the United States geography, history, and government. To make personnel selections, personal information on application forms and actual job performance during a six-month probationary period were used in addition to test performance (DuBois, 1970).

Following are the actual procedure, as specified by Dubois (1970), used by the U.S. Civil Service Commission to determine successful employees:

(1) The specific characteristics (e.g. level of intellectual functioning, interests, temperamental characteristics, etc) of individuals performing well in positions similar to the one(s) to be filled were identified.

(2) Examinations to assess these characteristics in applicants were developed.

(3) The examinations were administered in conditions that would not favor some candidates over others. For example, applicants were allowed the same amount of time to take the examinations.

(4) Examination numbers were used instead of names so that graders would not know whose examinations they were grading.

(5) A carefully-drawn point system was established for scoring purposes.

(6) Candidates were ranked in the order of their performance with the highest scorer standing the greatest chance of being hired (DuBois, 1970).

Though the United States and those European countries which adopted the Chinese test concept adapted it to their own needs and circumstances, they did retain some of its significant features. For instance, according to DuBois (1964), the idea of concealing the candidates' names from examiners was Chinese. Apparently, the Chinese even went a step further. They would get copyists to rewrite

the answers to avoid the candidates' handwriting giving away their identities. It was also the Chinese who came up with the idea of "uniform conditions" so that no candidate would be harmed or benefitted by differential circumstances under which the examinations were administered.

Ironically, it was just when the Chinese were abolishing their civil service examinations (1905) that the craze over them was developing elsewhere. Important to note, however, is the fact that it was not until 1916-1917 that psychologists started getting involved with vocational aptitude measurement in the United States (DuBois, 1970). In order to put the aptitude testing movement in the United States in perspective, the contributions of individuals such as Galton, Binet, Cattell, Jastrow, Goddard, and Terman must be examined. Though the work of these individuals concentrated on intelligence testing, their contributions to the development of aptitude testing is significant.

Frances Galton - Gould (1981) referred to British Sir Frances Galton (1822 - 1911) as the "apostle of quantification" (p.75). According to Gould, Galton was so fascinated by numbers that he believed he could measure anything - from the efficacy of prayer to degrees of boredom (Gould, 1981). DuBois (1970) referred to him as the "principal founder of the scientific study of human differences" (p.11). A cousin of Charles Darwin's, Galton held that individual differences were hereditary. Though he was aware that these differences could be used in employee selection, he was more interested in how they impacted human intelligence.

In 1884, Galton set up his Anthropometric Laboratory at the International Health Exhibition in order to collect data on variables he believed were related to intelligence. The anthropometric tests included bodily measurements, ability to perform some physical tasks, and reaction time. According to Sokal (1987), we owe our knowledge of statitical measurement and principles of normal distribution to Galton who used the data he collected to conduct correlation and regression studies. Altogether, Galton was able to collect data on 9,337 individuals with ages ranging form five to eighty. One American who was greatly impressed with Galton's work was James McKeen Cattell.

James Cattell - Cattell, a native Pennsylvanian born in 1860, graduated with honors from Lafayette College in 1880. Cattell traveled to Europe soon after graduation where he spent the winter of 1880-1881 studying at the University of Gottingen. Upon his return to the United States in 1882, Cattell was awarded a research fellowship by Johns Hopkins University. He worked in a physiological laboratory with a biology professor by the name of H. Newell Martin. However, he could not retain the fellowship through the following year. In 1883 the fellowship he had received was awarded to John Dewey. This was a result of disagreements Cattell had with the faculty and administration at Johns Hopkins. This prompted his decision to return to Europe to work toward a doctoral degree at the University of Leipzig. As an experimental psychology student, he conducted several reaction time experiments. Though Cattell by 1884 had read some of Galton's work, it was only later that the two

psychologists actually met and shared ideas on their common interest. Cattell was very impressed with the work Galton had done and was greatly influenced by him (Sokal, 1987).

On his return to the United States in 1889, Cattell was appointed professor of psychology at the University of Pennsylvania where he expanded on Galton's anthropemetry and called it mental testing. Unlike Galton, who was primarily interested in physiological measurements, Cattell included psychological characteristics in his measurements. He hoped his testing program would help him answer the following questions:

(1) To what extent are the several traits of body, of the senses, and of mind interdependent?

(2) How far can we predict one thing from our knowledge of another?

(3) What can we learn from the tests of elementary traits regarding the higher intellectual and emotional life (Sokal, 1987, p.33).

In 1891 Cattell transferred to Columbia College. There he expanded his tests to include craniometry, pitch perception, a wider variety of reaction time tests, and memory tasks. In 1894 he was granted permisssion to test every student admitted to Columbia. The analysis of the data he collected indicated that it was possible to assess an individual's ability to perform a number of specific tasks and thus quantify variability among individuals. Cattell, however, did not have any utilization plans for his findings.

It was not until late in the decade that his data was put to use

by one of his graduate assistants, Clark Wissler, who happened to be more mathematically inclined than Cattell. Wissler did correlations among the results of various tests and certain student variables. The results of the analysis were quite disappointing in that no correlations were found even between those tests that presumably tapped the same factor, and between any of the test results and student academic performance. It was this finding that led to the death of anthropometric mental testing as a movement in the United States (Sokal, 1987), and possibly gave way to aptitude testing as we know it today. Joseph Jastrow was another psychologist who made significant contributions in the area of aptitude testing by determining the factors that needed to be assessed by tests of aptitudes.

Joseph Jastrow - Jastrow earned his Ph.D. at Johns Hopkins University in 1886. Two years later he was appointed professor of psychology at the University of Wisconsin. He, like Cattell, was interested in anthropometry and had corresponded with Galton in 1887. Emulating Galton, he put up an anthropometric laboratory at the World's Columbian Exposition in Chicago in 1893. There he was able to test thousands of individuals using tests he had formulated. His tests were more like a cross between Cattell's and Galton's. This came as no surprise since Jastrow had been communicating with Galton and had served on an American Psychological Association (APA) committee charged with looking at the possibility of collaboration among all laboratories collecting mental and physical data when Cattel was APA president.

It was at a December 1897 APA meeting that Jastrow voiced his opinion about the areas that psychological tests should attempt to assess: the senses, motor capacities, and cognitive processes (Sokal, 1987). Looking at popular aptitude tests in use today (the GATB, for example), it would seem Jastrow's insights played a role in their formulation. Unfortunately, due to a nervous breakdown that Jastrow experienced in the late 1890's, he could not continue his work in assessment (Gould, 1981). Another psychologist who made a significant contribution to the area of testing was Henry Goddard.

Henry Goddard - Goddard earned his Ph.D. from Clark University and started teaching at a small state college soon afterwards. In 1906 he was appointed director of research at the New Jersey Training School for Feebleminded Boys and Girls at Vineland, NJ. For lack of better techniques and instruments, Goddard used anthropometric tests to assess the children's mental capabilities. Towards the end of the decade he traveled to France where he was introduced to one of the "gurus" in mental testing - Alfred Binet. Though the two psychologists were both interested in measuring individual differences, they differed in there approach. Binet's aim in testing was to identify those children that were not doing well in regular French schools so that appropriate remediation could be provided. Goddard, being a biological determinist, did not believe that education could do much for his charges, in fact he was more concerned about "morons" reproducing and propagating their kind (Gould, 1981).

Goddard's major contribution, however, was his bringing Binet's

tests to America, translating them into English, using them and documenting that they were much more effective than the anthropometric tests he had been using previously. As a result, Binet's (English version) tests finally replaced anthropometric mental testing (Sokal, 1987). Considering his contribution, it is probably fitting to provide a brief discussion on how Binet developed his tests and what they actually entailed.

Alfred Binet - As director of the psychology laboratory at the Sorbonne (in France), Binet's first attempt at studying intelligence was with the most popular means at the time--anthropometry. Nonetheless, as time went on he realized that physiological differences between the most able and the least able students in his studies were so minute that making decisions based on these could be erroneous. As a matter of fact, he found that in some cases poor students would fare better on some "crucial" physiological measurements than their more intelligent counterparts. Moreover, he found that his own biases could affect the measurements he recorded (Gould, 1981).

Therefore when Binet was asked by France's Minister of Public Education to construct some instrument that would identify marginal students for remediation purposes, he decided to turn to more "psychological" methods of measurement. The problem faced by Binet was the scarcity of information on the topic at the time. There were some records of Galton's work in physiological measurement and reaction time, but there were none on reasoning or any form of mental functioning.

To accomplish his assignment, Binet, working with a colleague, decided to develop a variety of tests aimed at assessing various aspects of mental functioning, particularly problem-solving, judgment (reasoning), and ability to learn. For example, one test required that students be able to detect inconsistencies, another to carry out commands, and another to infer reasons for specific occurences in a text. Yet another test similar to spatial perception subtests currently in use, required students to make drawings to show what pieces of paper cut from a folded diagram would look like reassembled (DuBois, 1970).

The tasks were graded according to their level of difficulty, and an age was assigned to each task. Age assignment depended on the lowest age at which "normal" children could complete the task. The mental age of a child was then determined by the highest level at which he/she could perform, and the general intellectual level was obtained by subtracting the mental age from the chronological age. It was William Stern, a German psychologist, however, who argued that dividing the mental age by the chronological age would be a better procedure for computing a child's intellectual level (Gould, 1981), a procedure still in use today.

Binet revised his test twice. The first version was out in 1905, the second in 1908, and the third just before his death in 1911. All three came with clear instructions on how to administer, score, and interpret them. Though Goddard was credited with introducing the Binet tests to America, it was Lewis Terman who came up with the revision of the Binet scale that has enjoyed the widest

acceptance in the United States: the Stanford-Binet (Minton, 1987).

Lewis Terman - As a doctoral student, Terman had shown interest in the types of mental processes involved in intelligence. For his dissertation he administered tests for "inventiveness and imagination, logical processing, mathematical ability, language, interpretation of fables, learning to play chess, memory, and motor ability...."(Minton, 1987). In his experiments he compared the performance of "bright" to that of the "not-so-bright" boys. He completed his doctoral work in 1905 (before the Binet scale was published).

In 1910, Terman got a position in the education department at Stanford University and was promoted to head of the department in 1922, a position that he held until his death. It was during his tenure at Stanford that he started working on the 1908 version of the Binet scale. He published his first revision in 1912 and the second - the renowned Stanford-Binet - in 1916. Terman was the first to include the Intelligence Quotient (IQ) in mental testing. He hoped that the Stanford-Binet would not be limited to assessing intellectual ability and predicting school performance, but would also be used in vocational assessment and counseling. (Minton, 1987). Nonetheless, despite all the work in intelligence testing, it was not until World War I that psychologists started applying themselves seriously to the systematic development of aptitude testing instruments.

Of course, the above assertion does not imply that no work had been done on aptitude testing before this time. In 1913, Hugo

Musternberg had shown that reaction time tests could predict job performance of switchboard operators and streetcar motormen (DuBois, 1970). Walter Dill Scott, Hugo's colleague at the Carnegie Institute, had also developed some tests that were useful in selecting successful salesmen. Hollingworth had written on the theory of vocatonal testing in 1915. Nonetheless, though all these endeavors were meaningful in their own right, the collaboration among psychologists of varied interests and backgrounds, and the rigor with which they worked to develop and administer aptitude tests during World War I was unparalled (DuBois, 1970). The most prominent figures in the movement (e.g. Goddard and Terman), have already been discussed, but there are others like Henry C. Link, Robert M. Yerkes, and Arthur S. Otis, who also contributed to the war effort and whose input will be discussed next.

Aptitude Testing and World War I

Robert Yerkes - The major advocate for aptitude testing during World War I was Robert M. Yerkes. As president of the APA, Yerkes saw the war as a great opportunity to establish psychology as a genuine science (Gould, 1981). When it became public that the United States was joining the war in 1916, Yerkes proposed that psychological testing (including aptitude testing) be part of military recruiting. A committee to develop and administer the tests was formed in May 1917. The membership included Bingham, Terman, and Goddard (DuBois, 1970). The committee decided that group tests would be more cost and time effective than individual tests. Fortunately, Terman's graduate student, Arthur S. Otis, had developed some group

tests and these served as sufficient models for the army tests (Minton, 1987). The committee worked hard assembling the tests and subtests and trying to establish their validity and reliability. Within four months, and not without disagreements among committee members, the Army Alpha, designed for literate recruits, was ready. The Army Beta was developed later for illiterate recruits and foreigners (Gould, 1981).

The Army Alpha consisted of the following eight, timed subtests:

1. Oral Directions
2. Arithmetic
3. Reasoning
4. Practical Judgment
5. Synonym - Antonym
6. Disarranged Sentences
7. Number Series Completion
8. Analogies and Information (DuBois, 1970).

More than 1.7 million recruits were administered the test. Each one was graded from "A" to "E", and recommendations for placement were provided. Usually one had to score at least a "C" to be recommended for an army officer position.

The Committee on the Classification of Personnel in the Army, with Bingham as chairperson, constructed and normed a number of trade tests. Information on the various trades was obtained from experienced mechanics, trade union officials, factory supervisors, the United States Bureau of Statistics, civil service examiners, and Army officers. The crucial requirements for each job were then

translated into test items, and expert judges were consulted to evaluate the test items before they were incoporated into the tests. By 1918, eighty-three oral trade tests, forty picture-form tests, and thirty performance tests had been constructed and standardized (DuBois, 1970).

In addition to the APA subcommittees, individuals continued to make contributions. Raymond Dodge, for instance, made an instrument that could be used both to train and select gun pointers in the navy. Thorndike also conducted several studies, one of which found a correlation between athletic ability and flying. Motor coordination and speed of response were also found to be good predictors of success in flying (DuBois, 1970).

Henry Link - In 1916 while working at an arms plant, Henry C. Link found that a card-sorting task could identify successful shell inspectors, whereas a tapping test could be used to select good gauges. He also did correlations on certain test results with success in various other occupational groups such as clerks, assemblers, typists, and production machine operators. Again, he found that the first step in the selection procedure was to do a job analysis - specifying the kinds of skills, knowledges, and personality traits required for successful performance; the second step was constucting and administering tests that would determine the degree to which those skills, knowledges, or personality traits were possessed; and the third step was correlating the results with later job success to distinguish between those instruments that were potentially efficient and those that were not (DuBois,1970).

Nonetheless, though the war provided an impetus for psychologists' involvement in aptitude testing, numerous other aptitude tests were later developed. Some of these, like musical and artistic aptitude tests, had no particular relevance to the war effort.

Early Tests of Musical and Artistic Aptitudes

Musical and artistic aptitudes cannot be ascertained by simply evaluating the product. Bingham (1937) asserted that the ability to judge the quality of others' work (aesthetic sensitivity) with a certain degree of precision was just as important.

A pioneer in musical aptitude measurement was Carl E. Seashore. Seashore's years of work on the psychology of music at the University of Iowa culminated in the identification of six proficiencies or subskills required for musical aptitude: "a sense of pitch, intensity, time, consonance, tonal memory, and rhythm" (DuBois, 1970, p.86). However, Bingham (1937) pointed out that how important each of these subskills is would depend on the musical activity in question. Detailed results of Seashore's work can be found in his book, Measures of Musical Talent, first published in 1919 and revised in 1939 (DuBois, 1970).

In collaboration with Meier, Seashore also developed the Meier-Seashore Art Judgement Test, an instrument intended to measure the "aesthetic sensitivity" mentioned earlier (DuBois, 1970). The test required the examinee to compare and judge two works of art that were identical except for one feature. Lewerenz and McAdory also identified the factors that "tests of ability and judgment in visual art " were supposed to tap (DuBois, 1970, p.87). According to

Bingham (1937), crucial attributes for artistic aptitudes include color sensitivity, a sense of balance, a sense of visual rhythm, and some form of motor coordination.

Clerical Aptitude Testing

Most of the ground work for clerical aptitude testing was done by Thurstone around 1919. He constructed and administered a standardized test of office work to over 5,000 clerks. The subtests included basic addition and subtraction, identifying misspelled words, code learning, name writing and alphabetizing, and matching proverbs. Thurstone found a .50 correlation between accuracy scores and actual job performance (DuBois, 1970).

Despite the fact that the clerical field is quite broad, Bingham (1937) identified four classes of abilities required of clerical workers in general: (a) perceptual abilities which involve the ability to observe words and numbers, (b) intellectual abilities which entail understanding meanings of words and symbols and decision-making skills, (c) numerical and verbal abilities, and finally, (d) motor skills including finger dexterity. According to Bingham (1937), tests of clerical aptitudes can be either specific, general, or a combination of both. Tests of general clerical aptitude are almost similar to intelligence tests - they include items that assess skills in all of the classes in the preceding paragraph. Specific tests, on the other hand, measure specific abilities required for specific clerical work like typing, proofreading, or accounting (Bingham, 1937).

Mechanical and Manual Aptitude Testing

DuBois (1970) pointed out that the distinction between mechanical and manual aptitudes is not that easy to pinpoint. This is because there is much overlap in the skills required to perform tasks that call for these kinds of abilities. Bingham (1937), however, made the following distinction: Whereas mechanically inclined individuals will often exhibit a sense of form, ability to visualize three-dimensional structure, ability to perceive space relations, and manipulative intelligence, manually-skilled individuals need only have good eye-hand coordination, finger dexterity, and speed and steadiness of movement. They need not understand the mechanical relationships underlying the operations.

W.F. Kemble developed a Manual Dexterity Pegboard Test in 1917. The test measured speed in picking up and assembing small parts. O'Connor later constructed a Finger Dexterity Test which used a metal plate and metal pins instead of a pegboard and wooden pegs (Bingham, 1937).

J.L. Stenquist was one of the earliest people to do work in mechanical testing. In 1915, he published the results of a study he had conducted using the Stenquist Construction Test/the Stenquist Test of Mechanical Ability (Bingham, 1937). These tests involved putting together real-life objects. Speed and accuracy were the major variables measured (Bingham, 1937). The test was also administered to more than 14,000 illiterate recruits during World War I (DuBois, 1970).

After the war, Stenquist continued his work on mechanical

aptitude testing. He developed two series of tests designed for elementary school children. He later developed two picture tests of mechanical aptitudes. Studies indicated a .88 correlation between the latter tests and the assembly tests he had developed earlier. A group of psychologists at the University of Minnesota revised the Stenquist test in the late 1920's and had its name changed to the Minnesota Assembly Test. Other improvements included the timing of the picture test which gave it more reliability (Bingham, 1937).

Spatial Aptitude Testing

According to DuBois (1970), tests of spatial aptitudes have been used to supplement measures of verbal intelligence in predicting success in fields such as architecture, physics, and engineering where the ability to perceive spatial relations and to understand the construction and operation of mechanisms is imperative. Some of the earliest tests of spatial aptitudes were the Minnesota Spatial Relations Test, the Kent Shakow Form Board, and the Yale Spatial Relations Test (Bingham, 1937, p. 172). The Army Beta, the non-language group test developed for illiterate recruits during World War I, did tap some spatial perception abilities. Among its subtests were a geometrical construction and a maze test, both of which DuBois (1970) asserts had a spatial component.

CHAPTER III

THE DEVELOPMENT OF MULTIPLE APTITUDE TESTS

The preceding chapter contained a brief description of how the U.S. Civil Service Commission developed its aptitude tests in the late 1800's. The chapter also included a discussion of the contributions of renowned psychologists and other scientists to the area of testing in general, and aptitude testing in particular, especially during the first World War. As stated previously, most of the aptitude tests constructed at the time were designed to assess aptitudes for specific occupational fields. However, the use of multiple aptitude tests has since become quite popular. Contained within this chapter will be an examination of the processes and procedures used in the development of multiple aptitude test batteries.

Multiple aptitude tests such as the General Aptitude Test Battery (GATB), the Differential Aptitude Test (DAT), and the Career Ability Placement Survey (CAPS), are all based on a conceptualization that mental abilities and aptitudes are multi-dimensional as opposed to unitary. For instance, based on Spearman's (1904) two-factor theory of mental abilities discussed in the first chapter, intercorrelations among ability tests are a consequence of a common or general ability factor 'g' which accounts for performance in all tests and an 's' factor which reflects competence in the specific area assessed by any one test or subtest.

Though Spearman failed to provide evidence for his claim, a number of British scholars did advance some mathematical strategies

which were able to show "that correlation matrices often contained evidence of group factors in certain sets of tests" (Carroll, 1982, p.69). Thurstone (1929), for example, produced evidence that verbal and numerical items often used in ability tests constituted "two relatively independent factors" (DuBois, 1970, p. 124). He demonstrated that though verbal and numerical tests had high correlations within themselves, correlations between the two types were not appreciable (Dubois, 1970). This suggested that performance in all verbal tests called for similar and perhaps related kinds of abilities--hence the intercorrelations.

Utilization of Factor Analysis

It was Thurstone (1935) who expanded Spearman's theory and came up with a technique that would identify the actual factors underlying the observed intercorrelations: factor analysis. Factor analysis is a statistical procedure employed to integrate or unify items that possibly share similar sources of variance into factors in order to achieve what Barclay (1990) refers to as "scientific parsimony" (p.204).

Thurstone (1935) did a factor analysis study involving fifty-seven mental test variables. The study yielded thirteen "centroid factors" which were further rotated to an even simpler structure of nine factors (Carroll, 1982, p. 70). According to Kaiser (1960) (cited in Carroll, 1982), more sophisticated analysis techniques have yielded more or less the same factors as Thurstone's initial study. Thurstone tested the factors when he constructed the Primary Mental Abilities Battery. The procedure enabled him to

identify at least seven independent factors underlying mental abilities. These were Numerical Ability, Verbal Ability, Spatial-Visualizing Ability, Memory, Induction (or generalizing) Ability, and Deduction (or reasoning) Ability (Barclay, 1990).

Thurstone's work is important because "it brought into general awareness the concept of differentiable mental ability factors...factors that can be shown to manifest themselves in performance significantly on more than one test variable" (Carroll, 19822, pp. 60. 70). It was this awareness that made it possible to construct versatile multiple aptitude tests that possessed potential for broad usage in educational, career, and rehabilitation counseling. It meant that no longer would it be necessary to have a separate test for each occupational area. Employers, school counselors, psychologists, and rehabilitation counselors could administer one battery and be able to make decisions regarding the individuals' potential for success in particular areas based on their general aptitudes.

The significance of Thurstone's (1935) work becomes evident in an examination of multiple aptitude instruments like the General Aptitude Test Battery, the Differential Aptitude Test, and the Career Ability Placement Survey, all of which sprang up after he published his work. The aptitudes these instruments attempt to assess are in most cases similar. For instance, all three batteries have one or more subtests measuring numerical ability, a test for verbal aptitude, and they all measure spatial perception (Prediger, 1989). But how do multiple aptitude test developers determine which items to

include in what subtest and what subtests converge on what factors? There are two procedures which serve this purpose: item analysis and factor analysis. A discussion on the development of the GATB will illustrate the process.

General Aptitude Test Battery (GATB)

The GATB was developed by the United States Employment Service (USES) shortly after World War II (DuBois, 1970). The major motive behind the construction of the GATB was to coordinate a battery of tests that would tap all the aptitudes that individual aptitude and intelligence tests attempt to measure. Items for inclusion in the GATB were selected from the previously used occupation specific tests. However, it was imperative to perform item-analysis to determine item validity (U.S. Department of Labor, 1970).

Item analysis involves administering a test to a preliminary group of respondents for purposes of identifying the "best" items. Three statistics are often used to determine the worth of each item: (a) the item discrimination index; (b) the percentage of respondents choosing each response to each item; and (c) item mean and standard deviation. The item discrimination index determines the discriminatory power of each item relative to the discriminatory power of the total test. An item which most high scorers get right and which most low scorers get wrong is supposedly good. The other two statistics focus on the options chosen. They show the extent to which each of the various options was used. Items on which respondents are spread are preferred over those on which they are clustered (Ary, Jacobs & Rozavich, 1985).

To gather item analysis data on the GATB, every test with the exception of the Motor Coordination test was administered to a sample of 200 participants. In order to ensure that validity information was gathered on each item, the tests were not timed; thus participants could attempt all the items. The number of correct responses to each item was used to determine the difficulty level of each item. The discriminating power was determined by the ratio of correct responses to each item in the highest percentile to correct responses in the lowest percentile. An item had to have a critical ratio of at least two for it to be included in the battery. Test-retest reliability was also established for the Motor Coordination test (U.S. Department of Labor, 1970).

Prior to the development of the GATB in the late 1940's, factor analysis studies had been conducted to identify the basic aptitudes that were assessed by the large number of aptitude tests then used by the U. S. Employment Service. However, as stated previously, individual tests could not be used as effectively in vocational counseling. They did not tap the wide spectrum of human aptitudes adequately (U.S. Department of Labor, 1970).

To identify the factors underlying the aptitudes, fifty-nine tests were used in several factor analysis studies. Of these, fifty-four had been developed by the U. S. Employment Service. Nine groups comprising a total of 2,156 men were used in the analysis studies. According to the U.S. Department of Labor, (1970), the primary goal was to construct a battery of tests that could be used in a variety of occupational situations without compromising their

validity. Tests measuring perceptual and spatial validity were emphasized though tests of verbal and numerical ability were also included.

The factors identified by the USES were Intelligence (I), Verbal Aptitude (V), Numerical Aptitude (N), Spatial Aptitude (S), Form Perception (P), Clerical Perception (Q), Aiming (A), Motor Speed (T), Finger Dexterity (F), Manual Dexterity (M), and Logic (L). However, further factor analysis for the ten groups suggested that Logic (L) was common to only two of the nine studies. As a result it did not qualify for inclusion in the battery.

Two criteria were used to determine which tests would be included: (a) internal or factorial validity, and (b) external or practical validity. Internal validity was determined by comparing factor loadings in the studies in which both the test and relevant factors occurred. To determine external validity, several occupational validation studies were conducted and relationships observed between each test and the obtained criteria served as the basis for validation (U.S. Department of Labor, 1970). Twelve tests that met the criteria were selected for inclusion in the GATB. These were the tests that had been proven to measure the nine aptitudes found to be critical in a wide variety of occupations: Intelligence or General Learning Ability, Verbal Aptitude, Numerical Aptitude, Spatial Aptitude, Form Perception, Clerical Perception, Motor Coordination, Finger Dexterity, and Manual Dexterity (U.S. Department of Labor, 1970).

CHAPTER IV

ISSUES INVOLVED IN USING APTITUDE TESTS TO PREDICT VOCATIONAL POTENTIAL

According to Mandeville and Brabham (1987), the rehabilitation process can be best summarized in the following steps: (a) selection and preliminary investigation; (b) accumulation of client evaluation data; (c) formulation of a vocational rehabilitation diagnosis (problem identification, eligibility declaration, selection of vocational objectives, identification of service needs); (d) planning, arranging and provision of services; (e) placement and case closure; and (f) evaluation of placement. Obviously, assessment is at the center of the initial steps in the rehabilitation process. Decision-making at every stage in the process is guided by assessment data. For instance, in order for a rehabilitation practitioner to establish whether a client is eligible for services, the nature of services most appropriate for the client, as well as the occupation in which the client can be the most productive and satisfied, assessment of client attributes is imperative. Therefore, the collection and application of assessment data is one of the most important roles of the rehabilitation practioner.

Frey (1984) stated, "All of what we do in rehabilitation depends upon our abilities to make appropriate, reliable, and valid assessments of those variables that facilitate the rehabilitation process. These assessments serve as a basis for all professional activity...." (p. 12). One set of attributes that rehabilitation professionals need to be aware of is client vocational aptitudes--the

extent to which the client can benefit from instruction of job-related skills; or to use the Department of Labor's (1972) definition, the extent to which the client possesses "the specific capacities or abilities required of an individual in order to facilitate the learning of some task or job duty" (DOL, 1972, p. 233).

Practical Utility of Aptitude Tests

This brings us to the next question: Do tests of aptitude have any practical utility? The definitions of aptitudes presented in the preceding sections indicate that aptitudes serve to predict potential for learning to perform certain job-related tasks (DOL, 1972; Warren, 1934; Woliver & Seaks, 1986). Whether aptitude tests really do have predictive validity has been a topic of discussion and has received criticism from numerous quarters (e.g. Halpern, 1984).

Halpern (1984) cited research which discounts the validity of the assumption that performance on aptitude tests can predict training outcomes or even actual job performance. According to Halpern, using a score on a "Manual Dexterity" subtest requiring an individual to place small pegs in holes to predict how fast and how well an individual can learn to change a tire, and making diagnosis, classification, and eligibility decisions on the basis of the individual's performance is somewhat unrealistic. What Halpern proposed instead was a system of measurement that would assess task performance directly "... within the context of its use" (p. 75).

Halpern's (1984) criticism seems to be similar to criticisms often advanced by behaviorists. According to Hunter (1986),

behaviorists view paper-and-pencil aptitude tests as inefficient measures of the potential to learn to perform a task. They see such tests as measures of academic learning. They advocate that a hands-on sample of performance be used to measure potential to perform on the job.

Another controversy surrounding aptitude testing involves whether it is specific aptitudes or general cognitive ability that should direct vocational rehabilitation counselor decisions (Gottfredson, 1986). It should be noted that the general cognitive ability (intelligence) score in most measures of aptitudes (e.g. the General Aptitude Test Battery) is often derived from combined scores of verbal aptitude, numerical aptitude, and spatial visualization. The crux of the controversy, therefore, becomes whether scores on these three subtests are an adequate indication of an individual's potential to perform any job.

There are a number of researchers who hold that the practical predictive validity of any test of aptitudes depends on its loading on the intelligence factor (Jensen, 1986; Hunter, 1986; Thorndike, 1986, Gottfredson, 1986). These researchers generally claim that specific aptitudes have an insignificant role in influencing an instruments's predictive validity.

According to Hunter (1986), whereas the predictive validity of aptitude tests is great for highly complex jobs, it is low for less complex jobs. Hunter further asserted that if specific aptitude tests had predictive validity, they would not predict job performance in other occupational areas. He claimed that the only specific

aptitude subtest that predicted performance better than other subtest was the clerical subtest on the Armed Services Vocational Aptitude Battery (ASVAB). Hunter suggested that this was the case because clerical tasks measured speed and other factors more than they did cognitive ability.

To support his argument, Hunter (1986) gave the example of a study conducted using the ASVAB in which a subtest designed to predict performance in mechanical work did not yield any better predictions than one designed to predict electronic and skilled work performance. To Hunter this meant that in both cases it was the general cognitive factor that influenced the score. Based on this finding, Hunter suggested that job learning predictions be made on the basis of general cognitive ability rather than "tailored aptitude composites". In fact, according to Hunter, general ability can go beyond job performance in its predictions; it may indicate creativity, problem-solving, and priority-setting, all of which are highly valued attributes in employees.

Though Crouse (1979) did not dispute that aptitude tests do predict satisfactory job performance, he contended that it was aptitudes that lead to the development of cognitive skills. This suggests that without an aptitude in an area, relevant cognitive skill development or expertise in that area would be unlikely. Therefore, according to Crouse, a test's predictive power depends on its accuracy in tapping aptitudes.

In response to the preceding assertions, Arvey (1986) pointed out that the probable reason for studies reviewed by the proponents

of general cognitive ability as a measure of job training and performance success is that these studies looked at "overall job performance" rather than at "narrow job performance" such as how a prospective employee "handles numerical calculations" (p. 416).

Humphreys (1986) did not seem convinced that general cognitive ability has better predictive validity than specific aptitude tests. According to Humphreys, one of the limitations of tests of general cognitive ability is that they do not give opportunities to exhibit acquired information and skills relevant to a specific job. Therefore, if they were used as a measure of how well clients could perform in certain positions, clients with specific skills that are not tapped by general ability tests might be left out unnecesarily. Humphreys' assertion was corroborated in a study by Chen and Rubin (1988) in which they developed prediction models for training program success involving clients with visual impairment. Univariate and multivariate analyses suggested that previous work experience was one of the most significant variables in predicting training success among this group of clients.

Though Linn (1986) did not dispute that tests of general cognitive ability have some degree of predictive validity for a wide range of jobs, he contended that tests of specific aptitudes had their place, especially if samples were relatively small. Also in agreement with Linn was Tyler (1986) who asserted that general cognitive ability measures alone did not account for all the variance among individuals; and in order for rehabilitation professionals to effectively help clients, they need to identify their special

talents--their specific aptitudes. Tyler criticised the research that supports the use of general ability scores to predict training and job success on the grounds that it is based almost exclusively on correlational studies, and ignores research in areas such as special education and rehabilitation which demonstrates that individuals can do productive work despite their relatively low cognitive ability level.

Another advocate for use of specific aptitude tests in rehabilitation and vocational counseling is Prediger (1989) who cited research that provides "evidence of work-related differences in the test-determined ability patterns of persons pursuing various occupations" (p. 3). For example, Prediger cited a follow-up study on Project TALENT participants to whom specific aptitude tests had been administered when they were still in high school. The follow-up study indicated that spatial visualization and mechanical reasoning aptitude subtests had accurately predicted successful auto mechanics, just as numerical computation subtests had correctly identified potential accountants.

Prediger (1989) concluded that "specialized abilities, taken together, are substantially more effective than 'g' in differentiating occupations" (p. 5). He concluded also that although general cognitive ability "is an important human trait, it does not produce flat profiles of specialized abilities--either of individuals or for occupational groups" (p. 6). In other words, general cognitive ability cannot adequately predict job performance because it is specialized abilities (or a composite of aptitudes) that distinguish among various job ability demands.

CHAPTER V

APPLICATION ISSUES INVOLVED IN TESTING PERSONS WITH DISABILITIES

It was stated earlier that aptitude tests are designed to make predictions about an individual's ability to learn to perform certain skills. The scores individuals make on these tests are often used to make decisions concerning their eligibility for admission into educational institutions or their fitness for placement in certain occupations. A review of the technical manuals of a typical aptitude test will reveal that nondisabled individuals were used to norm the test, and only a few are developed with the handicapped test taker in mind. As a result, the test taker with a disability is often at a disadvantage when having to take such tests. In addition to problems inherent in the disability, the individual has to face problems ranging form administrator bias to test characteristics which impede performance (U.S. Department of Labor, 1970; Pool, 1987).

Sherman and Robinson (1982) examined current practices in testing that pose barriers to the vocational evaluation process for persons with disabilities. They discovered that barriers still existed despite the stipulations of Section 504 of the Rehabilitation Act of 1973 which states that test scores of persons with disabilities should not reflect their disabilities but their aptitudes or abilities (Lincroft, 1984).

The purpose of this chapter is to discuss some issues involved in the vocational aptitude assessment of individuals with disabilities and to examine what is being done to ensure that their handicaps do not stand in the way of them entering the world of work

and fully participating in the mainstream of American society. Also, the validity of vocational aptitude tests when administered to persons with disabilities will be discussed.

Pool (1987) noted that most individuals, agencies, and institutions that are responsible for either assessing the aptitudes of the disabled or making decisions based on the assessments realized that aptitude tests can misrepresent the true potential of clients with disabilities. However, there was little agreement as to what could be done to remedy the situation. While some felt that modifications to the tests were imperative if fair estimates of capabilities were to be obtained, others felt that tests which are not designed for persons with disabilities should not be used to assess this population, or if they are used, they should be used without modifications. Of course, the issue of validity of such tests appeared to be of primary concern to these individuals.

Botterbusch and Michael (1985) did a detailed analysis of some tests currently employed in vocational evaluation. They looked at the purpose of each test, the test's component subtests, validity and reliability information, its characteristics, its usefulness in testing individuals with disabilities, and possible modifications to accomodate the various disabled groups. Their review included an analysis of the General Aptitude Test Battery (GATB). The GATB was normed on a sample of 4,000 workers from the general population. Reliability and validity studies of the GATB have been done using persons with disabilities. In one study, the dexterity tests (parts 9-12) of the GATB were administered to individuals with emotional

disturbances and a control group consisting of individuals without disabilities. On all parts, the controls scored higher than the individuals with emotional disabilities (The U.S. Department of Labor, 1970). This is consistent with other findings which indicate that the dexterity scores for persons with emotional disturbances tend to be lower than scores for persons without emotional disturbances (Floyd, 1964).

In a study involving persons with hearing-impairments, the results indicated that this group generally scored lower on the verbal ability subtest of the GATB (U.S. Department of Labor, 1970). Sherman and Robinson (1982) rightfully attributed the lower scores to the nature of their disability which interferes with their acquisition of verbal and communication skills. Since the verbal score is significant in determining the G (general intelligence and learning ability) score, persons with hearing impairment obtained lower G scores than the controls. Even more significant, however, is that when the USES Nonreading Measure of General Learning Ability (Nonreading G) score was substituted for the GATB G score, persons with deafness were found to qualify for more Occupational Aptitude Patterns (OAP's) (U.S. Department of Labor, 1970).

Despite the fact that the GATB has had the most extensive validation work and enjoys the widest usage by rehabilitation and employment agencies, it is not suited for use with the visually-impaired, according to Botterbusch and Michael (1985). It can, however, be administered to the hearing-impaired if an interpreter is available or written forms of the instructions are

provided. With regard to mental retardation, the degree of retardation will determine if an individual can take the test. Botterbusch and Michael recommended that some type of orientation or pretest be provided to enable the evaluator to determine the level at which the potential testee is functioning and then make the necessary accomodations. Botterbusch and Michael also urged evaluators to try to find a test that was specifically developed for individuals with disabilities and only resort to adapting tests designed for the general population if no tests exist for special needs clients.

The following is an integration of modification recommendations put forth by Sherman and Robinson (1982), Botterbusch and Michael (1985), and Pool (1987), for vocational aptitude test administration to persons with disabilities. There is a lot of overlap in the recommendations, and while some are more disability specific, others are general. The suggested adaptations for the various disability types involve test administration procedures, time limits, test content, and testing environment.

Visual Disabilities

As is the case with every disability, the degree of impairment is not uniform among this group. Consequently, whatever modifications are implemented should be commensurate with the severity of the handicap. The suggested modifications include using the large print or braille version of the test which may be obtainable from the American Printing House for the Blind; getting the client to repeat the instructions to the administrator to ensure that they are understood; presenting the material orally; and giving

a tape-recorded version of the test. To respond to the questions, the partially-sighted can use large print answer sheets, while those with a more severe impairment can use either a regular typewriter or a braille writer to answer the questions. Having the test taker give an oral response to the questions and the evaluator recording these on the appropriate answer sheets is another alternative.

Hearing Disabilities

Botterbusch and Michael (1985) suggested the use of pre-trials to familiarize the test taker with the format of the test and to enable the evaluator to determine the mode of communication (for example, lip-reading, signing, or finger-spelling) that the test taker normally uses. It is important for the evaluator to use the same communication mode because the client's understanding of the instructions depend on it. If the evaluator is unable to use the same mode of communication as the client, an interpreter should be used.

The hearing impaired client can be asked to repeat the questions to the administrator to ensure that he understands them. To start and end a testing session, lights can be flickered off and on. Demonstrations during practice should be exaggerated so that the test taker knows exactly what is expected. Video taped demonstrations can also be used.

Cognitive Deficiencies

Pre-trials with persons who have mental retardation will also help the evaluator determine the level of complexity that the client will find easiest to understand and then adjust the difficulty level

of the instructions as needed. Botterbusch and Michael (1985) also recommend the use of visual aids or other demonstrations to ascertain that the test taker understands the instructions. Having the test taker repeat the questions is again recommended. However, it might be necessary for the administrator to divide each set of instructions into subparts and make sure that the client understands each subpart before continuing to the next. Finally, the administrator should have the client repeat the whole set of instructions before going on with test administration. Because of their disability, clients with mental retardation might make mistakes in marking the answers. To guard against errors which can distort their true aptitude scores, the administrator can have persons with mental retardation give oral responses to the questions while the administrator records their answers for them.

Botterbusch and Michael (1985) also suggest the use of low literate and non-verbal forms of the tests whenever these are available. In one study involving individuals at different levels of intellectual functioning, none of the subjects with a G score of 75 or less were able to get a verbal aptitude score of 100 or more. This suggests a rather strong relationship between G scores and verbal aptitudes. Therefore, in order to minimize the effects of depressed verbal aptitudes on the total score, and to ensure that efficient assessment of other aptitudes occurs, the use of low literate and non-verbal tests as recommended by Botterbusch and Michael seems logical.

Physical Disabilities

Another area of concern in vocational aptitude testing is for persons with disabilities such as low back pain, carpal tunnel, or amputation. There are some obvious handicaps that disabilities of this type may pose during testing. The dexterity components of vocational aptitude tests require use of limbs and movements of the upper and lower body (e.g., eye-hand-foot coordination). If a client has an amputation or carpal tunnel, the evaluator may choose not to administer the dexterity components of the test. If the dexterity components are administered with modifications, the evaluator must note in the final report that the results were affected by the functional impairments of the disability. Similarly, a person with low back pain may need repeated breaks if the pain makes completing a vocational aptitude assessment difficult. Since vocational aptitude tests are timed, the results for a client experiencing pain during the tests will be affected. Again, the impact that the functional impairment of pain had on the results should be noted in the final report.

General Test Modifications

It should probably be stressed again that in cases where special tests for persons with disabilities are available, they should be used instead of the adapted versions of standard tests (Botterbusch and Michael, 1985). Time limits also become an issue with these adjustments. The modifications often demand that standard time limits be altered, usually extended. In the case of clients with visual impairment for example, reading braille requires more time

than reading regular print (Ragosta & Kaplan, 1986). According to Botterbusch and Michael, however, time extensions might not interfere with norms of power tests while they may distort the scores on speed tests.

The importance of ensuring that the test taker understands the instructions can not be overemphasized. Test content might need to be altered if the pertinent knowledge, skills or information are acquired through stimuli not accessible to persons with disabilities. Of course, a general sensitivity on the part of the administrator to the special needs of the clients and an awareness of the variability of clients within disability groups based on previous experiences and severity of the handicap would contribute greatly toward the just assessment of the client's aptitudes.

As stated previously, the major concern of the critics of modifying aptitude tests for administration to persons with disabilities are the effects these adjustments might have on the norms, the test material itself and, of course, the validity of the tests. Every rehabilitation practitioner realizes the importance of using an evaluation instrument that will accurately predict the performance of the client in a potential job situation (Pool, 1987). Rehabilitation counselors often find themselves in double bind situations in which they want to accomodate a client's handicap during testing and sometimes end up risking distorting the true picture of the client's aptitudes.

According to Sherman and Robinson (1982), not much validation research has been conducted on the modified versions of the tests.

The often cited reason is that it is difficult to get big enough samples to carry out such research. Sherman and Robinson recommended that researchers in different geographic areas work together to conduct such studies using their "pooled" samples. This would solve the problem of small samples and would ensure that the modified tests are standardized and validated.

Client Perspective

It is quite common for disability groups to have experts who advocate for their rights. Sometimes there are discrepancies between what these "experts" think is good for the disability groups and what members in these groups see as being good for themselves. In order to find out what individuals with disabilities thought about the modified testing arrangements provided them, Ragosta and Kaplan (1986) conducted a survey of individuals with disabilities who had taken modified versions of the Scholastic Aptitude Test and Graduate Record Examination.

Though the majority of the test takers reported being satisfied with the test adjustments, some concerns were raised with regard to the tests themselves, insufficient time, the presence of distractions in testing rooms, test administration, and administrator attitudes. For example, some reported being denied taking regular versions of the tests even if that is what they would have preferred. Some visually-impaired test takers complained that in translating the tests into other forms, the translators did not always make sure that the modified versions matched the standard versions. In some cases the braille, cassette, and printed versions of the tests did not

match, and other times the braille version of math tests did not include graphics that would have made it possible for the test takers to understand the concepts they needed to understand in order to answer the questions. In some cases where large print versions of the tests were given, the desks were not big enough. The value of the work done by Ragosta and Kaplan (1986) cannot be downplayed. What consumers have to say about these modifications as well as the results of norming and validation studies should be used in constucting the kinds of tests that will give the evaluators and rehabilitation counselors accurate data. Also, as Sherman and Robinson (1982) indicated, rehabilitation counselors should not put too much emphasis on test scores in making decisions about client counseling and placement. Tests are just "estimates" of an individual's potential. Use of work samples which are more congruent with job requirements could provide more accurate information with regard to a client's suitability for a particular job.

Testing, Orientation, and Work Evaluation in Rehabilitation (TOWER) and its shorter version MicroTOWER are systems with work samples that can be used to measure aptitudes of the clients. Though few, if any, independent validation studies have been done on these systems, they have been normed on a number of disability groups including clients with physical disabilities and mental retardation. Of course, expressed interests, work experience, social history, and educational backgound should not be discounted in making placement decisions (Sherman & Robinson, 1986).

CHAPTER VI

SUMMARY AND CONCLUSION

The first chapter of this monograph contained definitions of aptitude and presented a discussion of the relationship between aptitude and other constructs (intelligence and interest) equally as relevant to successful job placement and performance. The second chapter included a review of the history of testing and put aptitude testing in the broader context of the general testing movement. The contributions of the ancient Chinese, the United States Employment Service, and numerous associations and individuals from diverse persuasions and origins were examined. In the third chapter, an attempt to trace multiple aptitude test development using information provided by the United States Employment Service on its construction of the General Aptitude Test Battery was presented. In chapter four, there was a discussion of the practical utility of aptitudes and aptitude tests in vocational rehabilitation. Both sides of the controversy regarding the use of general cognitive ability measures as opposed to specific aptitude tests to collect data and make decisions pertinant to client rehabilitation and placement were also presented in chapter four. Finally in chapter five, the application issues involved in testing persons with disabilities were examined.

It would seem from the material presented in this monograph that despite the elusiveness of the construct of "aptitude," aptitude testing of vocational potential is necessary in order for rehabilitation counselors to make intelligent training and placement decisions. Bingham (1937) stated that there are three basic

assumptions on which the theory of aptitude and aptitude testing are based. First, individuals differ in terms of abilities. For instance, whereas some individuals exhibit excellent spatial perceptual skills, others have better abstract reasoning skills, and while some are endowed with both, others are not good in either. Second, an individual's abilities are not all equally good. An individual good in athletic ability might not be as good in abstract reasoning. Third, these differences within and between individuals are relatively enduring or stable (otherwise there would be no point in trying to measure them). A student who does well in third grade math is likely to do well in ninth grade algebra. In other words, present and past achievements are often indicative of future potential.

An objective of rehabilitation counseling is to assist individuals with disabilities understand their potentialities for employment. Brown (1970) defined a test as "a systematic procedure for measuring a sample of an individual's behavior under a given set of conditions or circumstances" (p.2). Aptitude tests, likewise, assess certain abilities of individuals as they exist at the time of testing, and the test results are used to estimate future potentialities. In order to appreciate the utility of vocational aptitude tests for making training and placement decisions, rehabilitation counseling students first need to understand what is meant by the construct vocational aptitude, also why and how tests are developed to assess it. It is proposed that this monograph will enhance students' understanding and appreciation of vocational

aptitude testing, thereby leading to effective and appropriate use of the tests when the students assume roles as rehabilitation professionals.

References

Allen, M. J. & Yen, W. M. (1979). Introduction to measurement theory. Monterey, CA: Brookes/Cole Publishing Co.

Arvey, R. D. (1986). General ability in employment: A discussion. Journal of Vocational Behavior, 29, 415-420.

Ary, D., Jacobs, L. C., & Rasavieh, A. (1985). Introduction to research in education. Chicago: Holt, Rinehart & Winston.

Ayers, G. E. (1969). Vocational rehabilitation counseling. In L. J. Zamir (Ed.), Dimensions in rehabilitation. Springfield, IL: Charles C. Thomas.

Barclay, J. R. (1990). Psychological assessment: A theory and systems approach. Malabu, FL: Krieger Publishing Company.

Bingham, W. V. (1937). Aptitudes and aptitude testing. New York: Harper & Brothers.

Botterbusch, K. F. & Michael, N. L. (1985). Testing and test modification in vocational evaluation. Menomonie, WI: Wisconsin University-Stout. (ERIC Document Reproduction Service No. ED 267 269)

Brookdale Community College (1984). Final report of research on vocational assessment of the physically handicapped. Lincroft, NJ: Brookdale Community College. (ERIC Document Reproduction Service No ED 262 209)

Brown, F. G. (1970). Principles of educational and psychological testing. Hinsdale, IL: The Dreyden Press.

Carroll, J. B. (1982). The measurement of intelligence. In R. J. Sternberg (Ed.), Handbook of Human Intelligence. Cambridge: Cambridge University Press.

Chen, S. C. & Rubin, S. C. (1988). The relationship of visually impaired client, personal characteristics and office skills training outcome. Journal of Applied Rehabilitation Counseling, 19(2), 3-6.

Crouse, J. (1979). The effects of academic ability. In C. Jencks et al. (Eds.), Who gets ahead? The determinants of economic success in America.

DuBois, P. H. (1970). A history of psychological testing. Boston: Allyn and Bacon, Inc.

Floyd, W. A. (1964). Aptitude testing with mental patients. Vocational Guidance Quarterly, 12, 203-206.

Frey, W. D. (1984). Functional assessment in the '80s. In A. S. Halpern & M. J. Fuhrer (Eds.), Functional assessment in rehabilitation. Baltimore: Paul H. Brookes.

Gottfredson, L. S. (1986). Societal consequences of the g factor in employment. Journal of Vocational Behavior, 29, 379-410.

Gould, S. J. (1981). The mismeasure of man. New York: Norton.

Guilford, J. P. (1967). The nature of human intelligence. New York: McGraw Hill.

Halpern, A. S. (1984). Functional assessment and mental retardation. In A. S. Halpern & M. J. Fuhrer (Eds.), Functional assessment in rehabilitation. Baltimore: Paul H. Brookes.

Humphreys, L. G. (1986). Commentary. Journal of Vocational Behavior, 29, (3), 421-437.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. Journal of Vocational Behavior, 29, 340-362.

Jensen, A. R. (1986). g: Artifact or reality. Journal of Vocational Behavior, 29, 301-331.

Linn, R. L (1986). Comments on the g factor in employment testing. Journal of Vocational Behavior, 29, 438-444.

Mandeville, K. & Brabham, R. (1987). The federal-state vocational rehabilitation program. In R. M. Parker (Ed.), Rehabilitation counseling: basics & beyond. Austin, TX: Allyn and Bacon, Inc.

Matarazzo, J. D. (1978). Wechsler's measurement and appraisal of adult intelligence (5th Ed.). New York: Oxford University Press.

McConnell, J. V. (1986). Understanding human behavior (5th Ed.). New York: Holt, Rinehart and Winston.

McGowan, J. F. (1969). Referral, evaluation, and treatment. In D. Malikin & H. Rusalem (Eds.), Vocational rehabilitation of the disabled: An overview. New York: New York University Press.

Minton, H. L. (1987). In search of the democratic ideal. In Sokal, M. M. (Ed.), Psychological testing and American society: 1890-1930. New Brunswick: Rutgers University Press.

Myers, G. E. (1941). Principles and techniques of vocational guidance. New York: McGraw Hill.

Pool, D. A. (1987). Psychological assessment of the disabled. Paper presented at the Annual Convention of the American Psychological Association, New York, Aug. 28 - Sept 1, 1987. (ERIC Document Reproduction Service No. ED 301 026).

Prediger, D. J. (1989). Ability differences across occupations: More than g. Journal of Vocational Behavior, 34, 1-27.

Ragosta, M. & Kaplan, B. A. (1986). A survey of handicapped students special test administrations of the SAT and GRE. Report No. 5. Studies of admissions testing and handicapped people. Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 275 697).

Schlenoff, D. (1983). Rehabilitation of unemployable mentally retarded persons: The avocational alternative. Rehabilitation Literature, 44(1-2), 27-28.

Sherman, S. W. & Robinson, N. M. (Eds.) (1982). Ability testing of handicapped people: Dilemma for Government, Science, and the Public. Washington, DC: National Academy of Sciences - National Research Council. (ERIC Document Reproduction Service No. ED 221 560)

Sokal, M. M. (1987). Psychological testing and American society: 1890-1930. New Brunswick: Rutgers University Press.

Spearman, C. (1927). The abilities of man. New York: MacMillan Company.

Sternberg, R. J. (1979). The nature of mental abilities. American Psychologist, 34, 214-230.

Thorndike, R. L. (1986). The role of general ability in prediction. Journal of Vocational Behavior, 29, 332-339.

Thurstone, L. L. (1935). The vectors of the mind. Chicago: University of Chicago Press.

Thurstone, L. L. (1938). Primary mental abilities. Chicago: University of Chicago Press.

Tyler, L. E. (1986). Back to Spearman. Journal of Vocational Behavior, 29, 445-450.

United States Department of Labor (1970). General Aptitude Test Battery (Manual). Washington, DC: U.S. Government Printing Office.

United States Department of Labor (1972). Handbook for analyzing jobs. Washington, DC: U.S. Government Printing Office.

Warren, H. C. (1934). Dictionary of Psychology. New York: Houghton Miffin Co.

Woliver, R. E. & Seaks, S. D. (1986). Intelligence and primary aptitudes. In R. B. Cattel & R. C. Johnson (Eds.), Functional psychological testing principles and instruments. New York: Brunner/Mazel.

Graduate Program in Rehabilitation Counseling
Graduate School/College of Education
124 Taylor Education Building
University of Kentucky
Lexington, Kentucky 40506
(606) 257-3834
FAX (606) 257-3835